

Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space

(fitness landscapes/neutral mutations/folding funnels/thermodynamic stability/lattice protein models)

ERICH BORNBERG-BAUER*[†] AND HUE SUN CHAN^{‡§}

*Theoretical Bioinformatics Group, German Cancer Research Centre Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany; and [‡]Department of Biochemistry, and Department of Medical Genetics and Microbiology, Faculty of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Edited by Peter G. Wolynes, University of Illinois at Urbana—Champaign, Urbana, IL, and approved July 26, 1999 (received for review February 19, 1999)

ABSTRACT Random mutations under neutral or near-neutral conditions are studied by considering plausible evolutionary trajectories on “neutral nets”—i.e., collections of sequences (genotypes) interconnected via single-point mutations encoding for the same ground-state structure (phenotype). We use simple exact lattice models for the mapping between sequence and conformational spaces. Densities of states based on model intrachain interactions are determined by exhaustive conformational enumeration. We compare results from two very different interaction schemes to ascertain robustness of the conclusions. In both models, sequences in a majority of neutral nets center around a single “prototype sequence” of maximum mutational stability, tolerating the largest number of neutral mutations. General analytical considerations show that these topologies by themselves lead to higher steady-state evolutionary populations at prototype sequences. On average, native thermodynamic stability increases toward a maximum at the prototype sequence, resulting in funnel-like arrangements of native stabilities in sequence space. These observations offer a unified perspective on sequence design, native stability, and mutational stability of proteins. These principles are generalizable from native stability to any measure of fitness provided that its variation with respect to mutations is essentially smooth.

The study of evolution requires an understanding of how sequences are mapped onto structures and functions. Fitness landscape is a useful conceptualization of sequence-space properties. It was originally proposed almost 70 years ago by Sewall Wright (1), who envisioned evolution as walks of populations on this landscape toward higher fitness.

Many evolutionary questions, such as those pertinent to random genetic drift as advocated by Motoo Kimura (2), entail modeling broad areas of the fitness landscape. Analytical models often assume a random mapping between genotype and phenotype, because the correlation among mutational effects proves to be mathematically complex to account for (3). However, such correlations are crucial in understanding neutral mutations and mutational stability. To address these issues, many recent theoretical efforts have been computational, focusing on constructing models of sequence-structure mapping for RNA (4–7) and proteins (8–22) that are motivated by various aspects of polymer physics. Because of the immense sizes of the systems, all these models involve significant simplifications (23–25). One of these models (8, 9) has been applied (10) to explore whether nonlethal mutations form a connected network, as envisioned by Maynard Smith (26).

Many proteins maintain their native structures while undergoing single and double mutations at many different sites. Of considerable evolutionary interest, therefore, is the number of converging sequences encoding for the same structure (9, 14, 16, 21). In a recent insightful study, Li *et al.* found that structures differ markedly in terms of their designability, i.e., their numbers of converging sequences, and that there are a small number of highly designable sequences (14).

Using a hydrophobic polar (HP) model with exhaustive conformational enumeration in two dimensions, one of us (16) recently obtained results consistent with that of Li *et al.* (17), including a Zipf-like distribution of designability. In addition, neutral nets are found to be centered around prototype sequences (16).

Another line of inquiry, beginning with the work of Bryngelson and Wolynes (27), has emphasized the importance of kinetic accessibility of the native structures as an evolutionary selection criterion. In this view, preferred structures are those that can be encoded by sequences with high native stabilities and minimal ruggedness on the folding landscape to allow for fast folding (13, 18–23). Structures preferred by the kinetic criterion are expected to be highly designable by the thermodynamic criterion of Li *et al.*, because they observed that high native stability is likely to correlate with high designability (14, 22).

Building on these results, we now turn our attention to sequence-space topologies of neutral nets (5, 16, 18), i.e., how sequences encoding for the same structure are related to one another by mutations. We observe the following features: (i) Independent of functional fitness, topology *per se* can lead to concentration of evolutionary population at some sequences. This perspective may rationalize certain mutagenesis experiments. (ii) The organization of native stabilities is funnel like among certain sets of sequences encoding for the same native structure. (iii) The presence of repulsive interactions can lead to more rugged sequence-space landscape.

Thermodynamic stability of the native structure is treated here as one possible “fitness” measure. Locally optimal sequences have previously been pictured as “peaks” (3) on the fitness landscape. Here we choose to conform to conventional imageries in physics and physical chemistry; we associate higher fitness with lower altitude on sequence-space landscapes instead. In this picture, neutral nets often are basins of attraction, with prototype sequences at their bottom.

MODELS OF NEUTRAL MUTATIONS

We study two models; both admit all possible permutations of two types of monomers along chain sequences, and chain

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office. Abbreviation: HP, hydrophobic polar.

[§]To whom reprint requests should be addressed. E-mail: chan@arrhenius.med.toronto.edu.

[†]Present address: Bioinformatics Group, The European Media Laboratory, D-69118 Heidelberg, Schloss Wolfsbrunnenweg 33, Germany.

conformations are configured on two-dimensional square lattices. We first consider the HP model (28), which assigns a favorable contact energy \mathcal{E} (<0) for each contact between two H monomers (an HH contact), whereas hydrophobic-polar (HP) and polar-polar (PP) contacts are neutral (zero contact energy). For comparison, we study also the "AB" model (29), with monomer types A and B. The contact energies for AA, BB, and AB contacts are, respectively \mathcal{E} , \mathcal{E} (<0), and $-\mathcal{E}$ (>0). The HP model is motivated by the physics of hydrophobic interactions. Native conformations in the HP model tend to have a hydrophobic core and a mostly polar surface, as in real proteins. On the other hand, the interactions in the AB model are very different; like monomers attract and unlike monomers repel. Hence the two types of monomers tend to segregate in native conformations, with mostly A monomers on one side and mostly B monomers on the other, which is not very protein like. The AB model is used here as a control, and also as a means to address effects of repulsive interactions and more disruptive mutations (29). Results below are presented for chain length $n = 18$.

For each of 2^{18} possible sequences in both models, exhaustive enumeration is used to identify the ground-state (lowest-energy) conformations among all 5,808,335 possibilities. There are, respectively, 6,349 and 34,700 sequences in the HP and AB models with a unique ground-state (native) conformation. They are used as model proteins (29). All single-point mutations (8, 11) among these sequences are determined. There are 16,340 such H \leftrightarrow P mutations (12) and 121,472 such A \leftrightarrow B mutations.

A neutral net is defined as a collection of unique sequence(s) encoding for the same native structure that are interconnected by single-point mutations. There are 1,706 and 16,270 neutral nets of various sizes in the HP and AB models, respectively. Fig. 1 shows the largest neutral nets. The Hamming distance between two sequences is the total number of monomers that are different along the alignment of the two sequences (16). Sequences within a neutral net that differ by only a single-point mutation (i.e., Hamming distance of one) are called neutral neighbors.

Most neutral nets center around prototype sequences (16) with the maximum number of neutral neighbors (Fig. 1). Motivated by their suggestive topologies, we first ask whether network connectivity alone can give rise to enhanced populations at the prototype sequences, based on an extremely simple model of evolutionary dynamics: for a neutral net with ω sequences, let the population of the i th sequence be P_i ($i = 1, 2, \dots, \omega$), and the mutation rate μ is the same for each of the n monomers. A mutation resulting in a sequence outside the neutral net is taken as lethal, corresponding to a population loss. Neglecting population entering the neutral net from the outside,

$$\frac{dP_i}{dt} = -\mu n P_i + \mu \sum_{j=1}^{A_i} P_{v_i(j)} \quad [1]$$

gives the time (t) dependence of the system, where $v_i(j)$'s label the A_i neutral neighbors of i . We note that the overall absolute population can increase or decrease depending on whether the reproductive rate [an additional term proportional to P_i in Eq. 1] is sufficient to offset losses to lethal mutations. However, this does not affect the steady-state relative population distribution, which is determined by the expected large- t behavior, $dP_i/(dt) = -\mu\lambda'P_i$, where λ' is some constant to be determined. Here μ can be factored out because the overall population decay rate in Eq. 1 must be proportional to the mutational rate. It follows that the μ -independent steady-state relative population distribution is given by the eigenvector for the largest λ in the eigenvalue problem

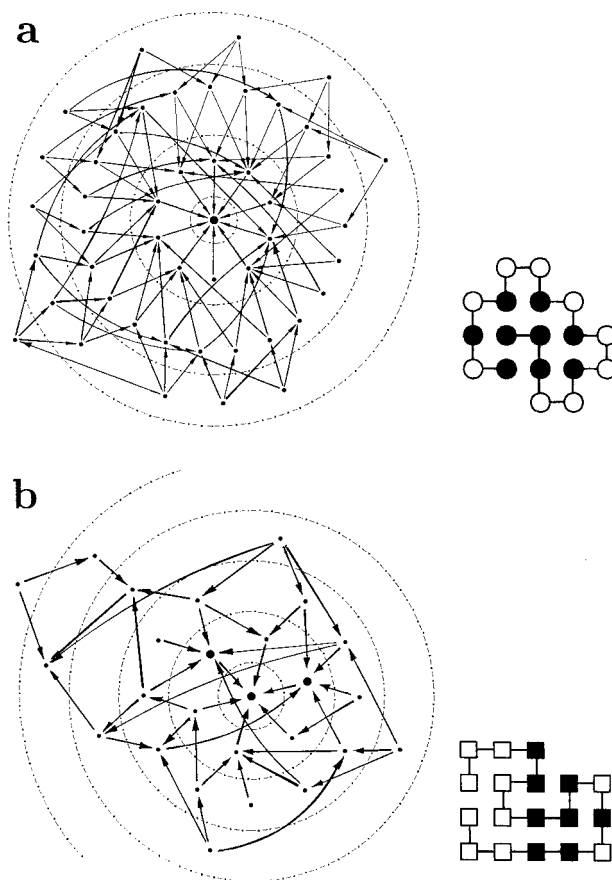


FIG. 1. Largest neutral net in (a) the HP model (48 sequences) and (b) AB model (26 sequences). The native structures are given in their respective prototype sequences. H, P, A, and B monomers are represented by filled and open circles and filled and open squares, respectively (29). The topology of each neutral net is shown by representing each sequence by a dot. A connecting line with arrow indicates that two sequences are neutral neighbors (16). Arrows point toward the sequence with higher native stability (see Fig. 2). Larger dots represent sequences with the maximum number of neutral neighbors within the neutral net. Concentric circles in dotted lines indicate Hamming distance from the prototype sequence.

$$\sum_{j=1}^{A_i} P_{v_i(j)} = \lambda P_i, \quad i = 1, 2, \dots, \omega, \quad [2]$$

where $\lambda \equiv (n - \lambda')$. It is straightforward to see that this leads to enhanced population at prototype sequences. If all ω populations are equal initially, mutations within the neutral net would not at first contribute to population change, because population flux from forward and backward mutations cancel. In this case, those sequences with fewer neutral neighbors will lose population faster because of their higher probabilities for lethal mutations. It follows that population distribution would shift subsequently in favor of sequences with more neutral neighbors.

For a more quantitative illustration, consider a hypothetical neutral net with perfect symmetry, which has two circles of sequences surrounding a single prototype sequence at the center. The prototype sequence has A_0 neutral neighbors. The A_0 sequences in the first circle each has A_1 neutral neighbors; they are also connected to a second circle of sequences, each of which has A_2 neutral neighbors. It can be shown that the ratios of steady-state populations between that of the prototype sequence (P_0) and one single sequence in the first (P_1) and the second (P_2) circles are given by $(P_0/P_1) = A_0/\sqrt{A_0 + A_2(A_1 - 1)}$ and $(P_1/P_2) = \sqrt{A_0 + A_2(A_1 - 1)}/A_2$.

For the special case of a symmetric neutral net with only one circle ($A_1 = 1, A_2 = 0$), these results reduce to $(P_0/P_1) = \sqrt{A_0}$, suggesting that in general steady-state population scales roughly as the square root of a sequence's number of neutral neighbors A .

For neutral nets in Fig. 1, steady-state populations are determined numerically. In both cases they peak at prototype sequences, with 6.31% for the HP and 7.09% for the AB neutral nets shown. In Fig. 1, the HP prototype sequence has 10 neutral neighbors; as for the rest, the highest and lowest steady-state populations are 4.07% and 0.97% for a sequence with 7 and 3 neutral neighbors, respectively. In the same figure, the AB prototype sequence has six neutral neighbors. Two other sequences have the same number of neutral neighbors; their steady-state populations are 6.91% and 6.89%. The lowest steady-state population is 1.04% for a sequence with two neutral neighbors.

These variations among steady-state populations are modest. This is a result of their roughly $\approx O(\sqrt{A})$ dependence. For sequences with two types of monomers, the maximum A is equal to the chain length n , whereas n is small for our highly simplified short-chain models. However, this general formulation suggests that the steady-state population distribution can be much more uneven and possibly highly peaked at prototype sequences for real proteins made up of 20 amino acid types with chain lengths $n \approx 100$, because A can then be of the order $19n$.

In this simple model of evolutionary dynamics, the fitness measure is effectively a step function—a constant favorable fitness for every sequence in the neutral net and lethal outside, and an effectively infinite population size is assumed. Realistically, other factors such as native stability (see below) and genetic drift because of finite population (30) are likely to skew this simple picture. The most important observation here, however, is that even with just the simple ingredients of neutral net topology and the existence of lethal mutants, uneven distributions that peak at prototype sequences can naturally arise, and this feature should be general because its derivation does not depend on any particular chain model.

FUNNELS IN SEQUENCE SPACE

In the present analysis, sequences within a neutral net are neutral only with respect to their ability to encode for the same native structure. We now consider the different thermodynamic stabilities of the native structure they encode. For each sequence, we enumerate the distribution of conformations over all possible energetic states. The density of states $g(E)$ is the number of conformations with energy E , where E is the sum of intrachain contact energies. The free energy of folding to one of the ground-state conformations is given by (9, 11)

$$\Delta G = E_N + k_B T \ln[\{g(E_N) - 1\}e^{-E_N/(k_B T)} + \sum_{E > E_N} g(E)e^{-E/(k_B T)}], \quad [3]$$

where $k_B T$ is Boltzmann constant times absolute temperature, E_N is the ground-state energy, and $g(E_N) = 1$ for the unique sequences considered in this section. A more negative ΔG means a more stable native structure. Here we use the “sticking” parameter $-\mathcal{E}/(k_B T)$ at the folding-denaturation midpoint ($\Delta G = 0$) as stability measure. In some neutral nets, there is more than one sequence with the maximum number of neutral neighbors. We then define the prototype sequence to be the one that *also* has the highest native stability.

Fig. 2 shows examples in which the most thermodynamically stable sequence is also the prototype sequence. A striking feature, especially for the HP case, is the funnel-like arrangements of the neutral-neighbor connections, which are the

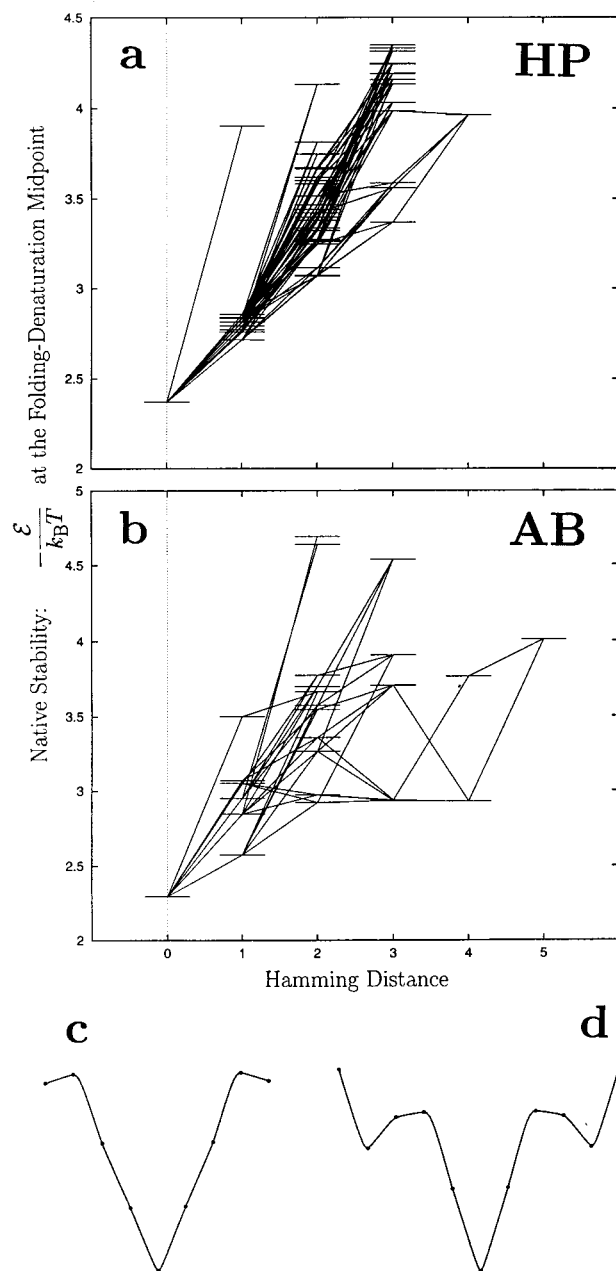


FIG. 2. Native stabilities of the sequences in the (a) HP and (b) AB neutral nets in Fig. 1 are represented as horizontal lines. The horizontal axis indicates Hamming distance from the prototype sequence. Neutral mutations are indicated by lines connecting horizontal levels. Heuristic views of (c) the HP and (d) AB “superfunnels” are traces through average stabilities of the sequences as a function of Hamming distance (dots). In c and d, the bottom at the center of each funnel corresponds to the prototype sequence; horizontal displacement from the center in either direction corresponds to increasing Hamming distance from the prototype sequence.

“kinetic adjacencies” (12) in evolution. On average, native stability decreases for sequences further away from the prototype sequence (Fig. 2 c and d). This appears to be a general feature of sequence space, irrespective of whether native stability *per se* is favored by evolutionary selection. Because each sequence is itself associated with a presumably funnel-like folding landscape consisting of all conformations (23, 24, 31), sequence-space funnels represent a higher level of organization in a “cross product” of the sequence and conformational spaces, while sharing qualitatively similar features with conformational-space funnels (31). Hence we propose adding the

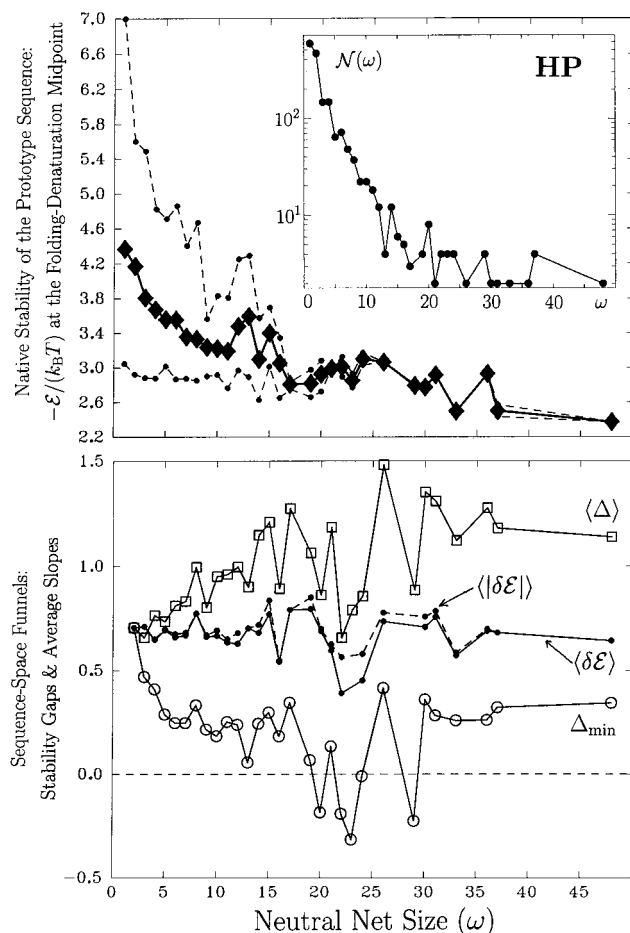


FIG. 3. Superfunnel geometry. (Upper) Thermodynamic stability of prototype sequences. For neutral nets of a given size, diamond shows the average stability, whereas dots show the maximum and minimum stabilities among the prototype sequences from different neutral nets. (Inset) $N(\omega)$ is the number of neutral nets with size ω . All solid or dashed lines linking data points in Figs. 3 and 4 serve merely as visual guides. (Lower) For a given neutral net, Δ is the difference in thermodynamic stability [measured in transition-midpoint $-\mathcal{E}/(k_B T)$] between a nonprototype sequence and the prototype sequence. The average stability gap $\langle \Delta \rangle$ is the average of Δ over all nonprototype sequences in the net; thus it provides a measure of “depth” of a neutral net. The minimum stability gap Δ_{\min} is the smallest value of Δ within a neutral net. Hence $\Delta_{\min} < 0$ implies the neutral net is not a superfunnel. Averages of $\langle \Delta \rangle$ (squares) and Δ_{\min} (circles) over neutral nets of given sizes are plotted. For neutral nets that satisfy the superfunnel criterion ($\Delta_{\min} \geq 0$), average slopes are also computed. For every neutral mutation, $\delta \mathcal{E}$ is a sequence-space slope. It is equal to the transition-midpoint $-\mathcal{E}/(k_B T)$ of the sequence one Hamming step further from the prototype sequence minus that of its neutral neighbor that is one step closer. $\langle \delta \mathcal{E} \rangle$ is the average of $\delta \mathcal{E}$ over all mutations within a neutral net. Because some neutral mutations lead to negative slopes ($\delta \mathcal{E} < 0$), we also compute the average of their absolute values, $\langle |\delta \mathcal{E}| \rangle$. Averages of these two quantities over neutral nets of given sizes are plotted as dots connected by solid ($\langle \delta \mathcal{E} \rangle$) and dashed ($\langle |\delta \mathcal{E}| \rangle$) lines.

prefix “super” to their description. We demonstrated above that prototype sequences are intrinsically favored even in the absence of any functional or reproductive advantage. If higher native stability is correlated with enhanced fitness, which may be a reasonable assumption for some biological activities, the existence of superfunnels in sequence space would imply an even higher concentration of steady-state evolutionary populations at the prototype sequences.

Thermodynamic statistics of all neutral nets in the two models are given in Figs. 3 and 4 as functions of neutral net

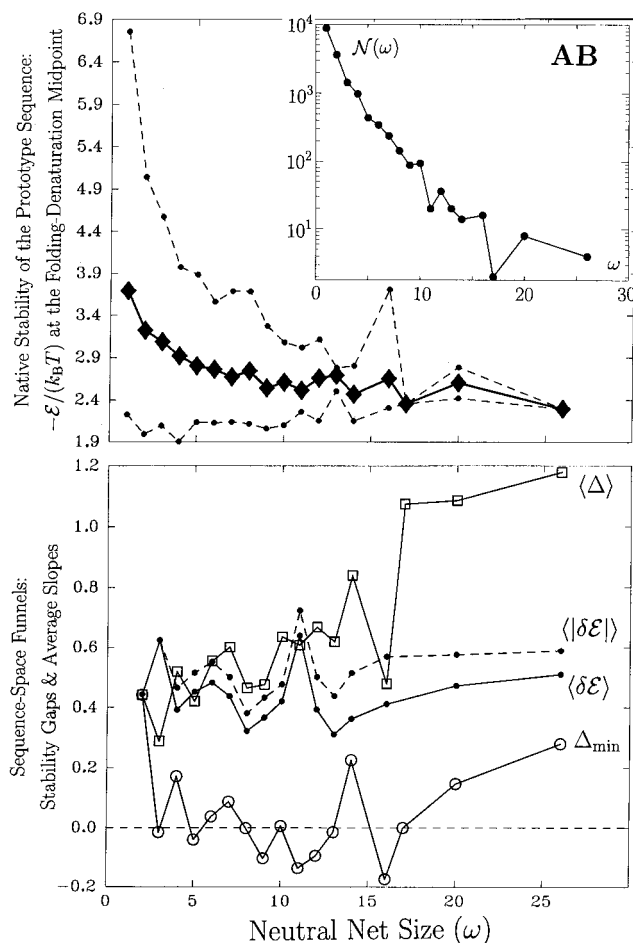


FIG. 4. Same as Fig. 3 except for AB sequences.

size. Figs. 3 (Lower) and 4 (Lower) show that a majority of neutral nets conform to the superfunnel paradigm. There are exceptions: in some neutral nets, sequences with the maximum number of neutral neighbors do not have the highest thermodynamic stability (see Δ_{\min} traces). These cases constitute only a minority. In the HP model, this occurs in 75 neutral nets, which comprise 11.2% of the 668 neutral nets with more than two sequences. Collectively, they contain 748 sequences, which is 11.8% of all unique sequences. In the AB model, 1,348 neutral nets do not conform to the superfunnel paradigm, which is 35.5% of the 3,882 nets with more than two sequences, and they involve 6,484 sequences, 18.7% of all unique sequences. In these situations, the dominant population would be determined by two competing evolutionary effects—the selective advantage of native thermodynamic stability vs. the neutral net topology effect described above.

For the majority of neutral nets that are superfunnels, there is a clear correlation between neutral net size and the stability of the prototype sequence (Figs. 3 and 4 Upper). This observation is consistent with the conclusions of Li *et al.* (14) and Melin *et al.* (22) on proteins and of Wuchty *et al.* (7) on RNA. This can also be seen here from the fact that, on average, the depth of superfunnels increases with size (Figs. 3 and 4 Lower, $\langle \Delta \rangle$ traces). The average slope of superfunnels $\langle \delta \mathcal{E} \rangle$, however, does not show any significant systematic increase or decrease with size.

These general trends are observed in both the HP and AB models, but details of the superfunnels depend on the intra-chain interactions of the sequences. Because of the repulsive interactions, mutations in the AB model are more disruptive. As a result, the AB sequence space is more fragmented, and

its neutral nets are on average smaller than that in the HP sequence space (Figs. 3 and 4 *Upper*). The largest HP neutral net coincides with the largest set (neutral set) of converging sequences (9, 16) encoding for the same structure. For the AB model, the structure encoded by the largest neutral net is not identical to that with the largest neutral set. The latter has 76 encoding sequences, but it is fragmented into 14 neutral nets.

HP superfunnels are also smoother. For instance, 98 of the 99 mutational connections in the largest HP neutral net in Figs. 1 and 2 have “positive slopes,” i.e., they are directed toward sequences with higher stabilities as they approach the prototype sequence. Only one mutational connection has a “negative slope.” There is no evolutionary “kinetic trap” on this HP superfunnel, because every nonprototype sequence has at least one neutral neighbor with higher thermodynamic stability [smaller $-E/(k_B T)$ at $\Delta G = 0$]. The AB neutral net in the same figures are more rugged in that 9 of the 46 mutational connections have negative slopes. The heuristic drawing in Fig. 2d also suggests that there are evolutionary “kinetic traps” in this AB neutral net. Indeed, there is one trap at Hamming distance 4, but it is very shallow. AB superfunnels are more rugged in general. This is illustrated by the much larger discrepancies between the $\langle \delta E \rangle$ and $\langle |\delta E| \rangle$ traces in Fig. 4 (*Lower*) vs. that in Fig. 3; $\langle |\delta E| \rangle - \langle \delta E \rangle$ describes the prevalence of negative slopes and is therefore a measure of superfunnel ruggedness.

MULTIPLY-DEGENERATE SEQUENCES

So far, we have assumed that only unique sequences are viable. However, it is conceivable that a sequence with more than one ground-state conformation [i.e., degeneracy $g \equiv g(E_N) > 1$] (29) can still possess the function performed by any one of its ground-state conformations. As a first approximation, we may assume that the activity specific to a given conformation is proportional to the fractional population p of that conformation. Its free energy of folding $\Delta G = -k_B T \ln[p/(1-p)]$ at a given intrachain sticking can then be used to characterize this activity. A smaller ΔG implies a higher stability for the

functional form. ΔG is calculated from the sequence’s density of state by using Eq. 3. Because multiply-degenerate ($g > 1$) sequences can never attain more than $p = 50\%$ population for any one of its g ground-state conformations, their ΔG s are always positive.

By including sequences with as many as six ground-state conformations, an *extended* HP neutral net for the one in Figs. 1 and 2a is constructed (Fig. 5). Every $g > 1$ sequence in this net has the HP structure in Fig. 1 as one of its ground-state conformations. Fig. 5 provides these sequences’ free energies of folding to this structure, resulting in a larger superfunnel with the same general features as that in Fig. 2a. The maximum Hamming distance remains unchanged. On average, the $g > 1$ sequences are further away from the prototype sequence than the unique ($g = 1$) sequences (average Hamming distance of 2.6 vs. 2.1).

These observations suggest a generalization of the superfunnel concept. Basins of attraction in sequence space can be substantially enlarged to encompass more sequences if multiply-degenerate sequences are to some degree viable. In this scenario, some of the $g > 1$ sequences that have other encodable conformation(s) (29) in their ground states can serve as “switches” (16) to facilitate evolution to other neutral nets. In Fig. 5, 13 of the $g = 2$ sequences share this property.

GENERALIZATIONS AND DISCUSSION

Our model results strongly suggest that “plasticity” or mutational stability of a sequence is correlated with its thermodynamic stability. We believe that this general conclusion follows directly from a fundamental principle of sequence design—that it is important to both design in the target structure and *design out* nontarget structures (32). Thus native states of better designed sequences are energetically more separated from their nonnative conformations, implying that they have higher thermodynamic stabilities (14, 22). Some threshold native stability may also be needed to avoid misfolding on multimerization and aggregation (33). Insofar as stability of a given native structure varies relatively smoothly in sequence space, a superfunnel-like organization is likely. In evolutionary terms, this means that in most cases the wild-type sequence may be identified with the prototype sequence, and that most if not all single-point mutations on the wild-type sequence would be thermodynamically destabilizing. In light of these considerations, the generality of our conclusion may transcend the two models studied here, and may also be independent of questions regarding what model contact interactions and chain representations are more protein like (21, 29).

Following this logic, it appears that any fitness measure could lead to superfunnel-like organizations of that measure, provided that its variation is relatively smooth in sequence space. Several recent evolutionary studies are based on assumed selective advantages for sequences having fast folding kinetics in addition to thermodynamically stable native structures (13, 18–22). We note that mutational effects on model folding kinetics can sometimes be subtle (12, 34), and actual dynamic simulations have shown that some kinetic properties cannot always be reliably derived from the density of states alone (35) as used in some studies (18, 21). A recent experiment also shows that mutations on wild type are most likely destabilizing, but their kinetic effects are less predictable (36). Nevertheless, inasmuch as these “foldability” criteria do vary smoothly (18), superfunnel-like organization of foldability is expected. Indeed, a funnel-like variation of a “frustration” measure of kinetic accessibility along a sequence similarity parameter has recently been reported in an off-lattice study by Nelson and Onuchic (20).

Our goal here is to establish a conceptual framework. Many subtleties and complexities of biological structure and function (37) are neglected. For instance, it has been suggested that too

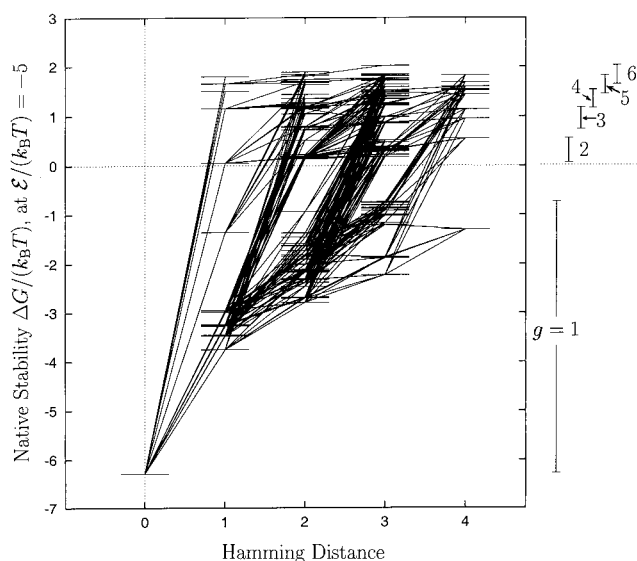


FIG. 5. Extended neutral net for the HP structure in Fig. 1, with 146 sequences. Same as Fig. 2a, except that native stabilities of the sequences are now measured by free energy of folding (Eq. 3) and that multiply-degenerate sequences (degeneracy $g \leq 6$) are included. Numbers of sequences with $g = 1, 2, 3, 4, 5$, and 6 in this net are 48, 22, 16, 27, 14, and 19, respectively. The vertical bars and numbers on the right indicate the range of stability levels for sequences with different g s.

much thermodynamic stability and conformational rigidity can be detrimental to function (38) (see also ref. 39). Real functional fitness is not expected to always correlate with native stability. The application of our results may also be less straightforward if the functional form of a protein is a multimer instead of a single-chain monomer (40). These limitations notwithstanding, the superfunnel scenario appears to be in general qualitative agreement with experiments. In an extensive mutagenesis study of 290 single-point mutations on the wild-type sequence of staphylococcal nuclease, only 33 lead to relatively small thermodynamic stabilization. All of the rest are destabilizing to various degrees (41–43). Consistent with the ruggedness consideration above, sets of mutations that are energetically more disruptive lead to higher probabilities of mutant stabilization: 2/83, 11/103, and 20/104 of the mutations on large hydrophobic (41), polar and uncharged (42) and ionizable (43) amino acid residues, respectively, lead to mutants more stable than the wild type.

A noteworthy finding here is that neutral net topology *per se* can be an important determining factor of evolutionary population. In some cases, mutations on wild-type sequences are found to be both stabilizing and function enhancing (44), which is puzzling from an evolutionary perspective that focuses exclusively on functional fitness (45). However, this may be rationalizable without invoking unspecified biological complexities if these mutants turn out to be mutationally unstable themselves (i.e., have few viable neutral neighbors). Our results show that it is possible for neutral net topology and functional fitness (such as native stability in our models) to have opposing evolutionary effects on population distribution. This hypothesis should be testable by experimental mapping of neutral net topologies of real proteins.

We thank José Onuchic and two anonymous referees for helpful comments on the manuscript. H. S. C. thanks the Medical Research Council of Canada (MT-15323) and the Connaught Fund for financial support.

- Wright, S. (1932) in *Proceedings of the Sixth International Congress on Genetics, Vol. 1*, ed. Jones, D. F. (Brooklyn Botanic Gardens, New York), pp. 356–366.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge).
- Kauffman, S. & Levin, S. (1987) *J. Theor. Biol.* **128**, 11–45.
- Fontana, W. & Schuster, P. (1987) *Biophys. Chem.* **26**, 123–147.
- Huynen, M. A., Stadler, P. F. & Fontana, W. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 397–401.
- Fontana, W. & Schuster, P. (1998) *Science* **280**, 1451–1455.
- Wuchty, S., Fontana, W., Hofacker, I. L. & Schuster, P. (1999) *Biopolymers* **49**, 145–165.
- Lau, K. F. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 638–642.
- Chan, H. S. & Dill, K. A. (1991) *J. Chem. Phys.* **95**, 3775–3787.
- Lipman, D. J. & Wilbur, W. J. (1991) *Proc. R. Soc. London Ser. B* **245**, 7–11.
- Shortle, D., Chan, H. S. & Dill, K. A. (1992) *Protein Sci.* **1**, 201–215.
- Chan, H. S. & Dill, K. A. (1994) *J. Chem. Phys.* **100**, 9238–9257.
- Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 1282–1286.
- Li, H., Helling, R., Tang, C. & Wingreen, N. (1996) *Science* **273**, 666–669.
- Bussemaker, H. J., Thirumalai, D. & Bhattacharjee, J. K. (1997) *Phys. Rev. Lett.* **79**, 3530–3533.
- Bornberg-Bauer, E. (1997) *Biophys. J.* **73**, 2393–2403.
- Ebeling, M. & Nadler, W. (1997) *Biopolymers* **41**, 165–180.
- Govindarajan, S. & Goldstein, R. A. (1997) *Biopolymers* **42**, 427–438.
- Saito, S., Sasai, M. & Yomo, T. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 11324–11328.
- Nelson, E. D. & Onuchic, J. N. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 10682–10686.
- Buchler, N. E. G. & Goldstein, R. A. (1999) *Proteins Struct. Funct. Genet.* **34**, 113–124.
- Melin, R., Li, H., Wingreen, N. S. & Tang, C. (1999) *J. Chem. Phys.* **110**, 1252–1262.
- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins Struct. Funct. Genet.* **21**, 167–195.
- Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4**, 10–19.
- Thirumalai, D. & Woodson, S. A. (1996) *Acc. Chem. Res.* **29**, 433–439.
- Maynard Smith, J. (1970) *Nature (London)* **225**, 563–564.
- Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528.
- Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995) *Protein Sci.* **4**, 561–602.
- Chan, H. S. & Dill, K. A. (1996) *Proteins Struct. Funct. Genet.* **24**, 335–344.
- Futuyma, D. J. (1998) *Evolutionary Biology* (Sinauer, Sunderland, MA), 3rd Ed., chapter 11, pp. 297–335.
- Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725.
- Yue, K. & Dill, K. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4163–4167.
- Harrison, P. M., Chan, H. S., Prusiner, S. B. & Cohen, F. E. (1999) *J. Mol. Biol.* **286**, 593–606.
- Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1998) *Fold. Des.* **3**, 183–194.
- Chan, H. S. & Dill, K. A. (1998) *Proteins Struct. Funct. Genet.* **30**, 2–33.
- Kim, D. E., Gu, H. & Baker, D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 4982–4986.
- Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990) *Science* **247**, 1306–1310.
- Rasmussen, B. J., Stock, A. M., Ringe, D. & Petsko, G. A. (1992) *Nature (London)* **357**, 423–424.
- Daniel, R. M., Smith, J. C., Ferrand, M., Héry, S., Dunn, R. & Finney, J. L. (1998) *Biophys. J.* **75**, 2504–2507.
- Milla, M. E., Brown, B. M. & Sauer, R. T. (1994) *Nat. Struct. Biol.* **1**, 518–523.
- Shortle, D., Stites, W. E. & Meeker, A. K. (1990) *Biochemistry* **29**, 8033–8041.
- Green, S. M., Meeker, A. K. & Shortle, D. (1992) *Biochemistry* **31**, 5717–5728.
- Meeker, A. K., Garcia-Moreno, B. E. & Shortle, D. (1996) *Biochemistry* **35**, 6443–6449.
- Hecht, M. H., Hehir, K. M., Nelson, H. C. M., Sturtevant, J. M. & Sauer, R. T. (1985) *J. Cell. Biochem.* **29**, 217–224.
- Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. (1986) *Proteins Struct. Funct. Genet.* **1**, 43–46.